TEC-0061

Vision-Based Navigation and Recognition - Second Annual Report

Ariel Rosenfeld

Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

June 1995

Prepared for:
Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714

Monitored by:
U.S. Army Corps of Engineers
Topographic Engineering Center
7701 Telegraph Road
Alexandria, Virginia 22315-3864

US Army Corps of Engineers
Topographic Engineering Center

T
E
C

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1995	Second Annual Report Apr. 1993 - Mar. 1994

4. TITLE AND SUBTITLE

Vision-Based Navigation and Recognition - Second Annual Report

5. FUNDING NUMBERS

DACA76-92-C-0009

6. AUTHOR(S)

Azriel Rosenfeld

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Advanced Research Projects Agency
3701 North Fairfax Drive, Arlington, VA 22203-1714

U.S. Army Topographic Engineering Center
7701 Telegraph Road., Alexandria, VA 22315-3864

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

TEC-0061

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report summarizes image understanding research dealing with many aspects of both navigation and recognition. This research has dealt with the following ten areas: (a) Parallel algorithms, (b) Invariant properties, (c) Image registration, (d) 3-D recovery, (e) Motion analysis, (f) Vision-based navigation, (g) Function-based recognition, (h) Face recognition, and (i) Document understanding. The work done in these areas is summarized in Sections 2-10 of this report. Further details about this work can be found in twelve technical reports issued on the Contract during the period April 1993 - March 1994. A bibliography of these reports is given in Section 11 of this report; the numbers in brackets in Sections 2-10 refer to this list.

14. SUBJECT TERMS

Parallel algorithms, 3-D recovery, motion analysis, vision-based navigation, Object recognition

15. NUMBER OF PAGES

27

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

Contents

List of Figures

PREFACE

This research is sponsored by the Advanced Research Projects Agency (ARPA) and monitored by the U.S. Army Topographic Engineering Center (TEC) under Contract DACA76-92-C-0009, titled "Vision-Based Navigation and Recognition - Second Annual Report." The ARPA Program Manager is Dr. Oscar Firschein, and the TEC Contracting Officer's Representative is Ms. Lauretta Williams.

1 Introduction

Image understanding research at the Center for Automation Research of the University of Maryland at College Park deals with many aspects of both navigation and recognition. This report summarizes the research conducted under Contract DACA76-92-C-0009 (ARPA Order 8459) during the period April 1993 – March 1994.

The research conducted under the Contract has dealt with ten areas:

(a) Parallel algorithms,

(b) Invariant properties,

(c) Image registration,

(d) 3-D recovery,

(e) Motion analysis,

(f) Vision-based navigation,

(g) Function-based recognition,

(h) Face recognition, and

(i) Document understanding.

The work done in these areas is summarized in Sections 2–10 of this report. Further details about this work can be found in twelve technical reports issued on the Contract during the period April 1993 – March 1994. A bibliography of these reports is given in Section 11 of this report; the numbers in brackets in Sections 2–10 refer to this list.

2 Parallel algorithms [7]

The following research has dealt with algorithms for parallel processing of geometric structures, with emphasis on algorithms for visibility and triangulation problems. A number of parallel algorithms have been developed and evaluated using a mixture of theoretical and practical criteria. Most of the algorithms have been implemented on parallel machines such as the Connection Machine, and experimental results have been obtained and analyzed.

In connection with the problem of determining visibility on digital terrain models, an algorithm has been developed for computing point-to-region visibility using propagations between neighboring terrain cells, as well as a ray-structure-based algorithm for region-to-region visibility. In particular, a parallel algorithm has been designed for computing visibility on polyhedral terrain models. Its complexity is $O(\log^2 n)$ time and $O((n\alpha(n) + k)\log n)$ operations on a CREW PRAM, where n and k are the input and output sizes respectively, and α is the inverse of Ackermann's function. The vertex-ray method for computing approximate visibility on polyhedral terrain has also been studied because of its suitability for data-parallel implementation.

A method of triangulating a terrain surface in parallel has been designed. The method is able to construct either a hierarchical triangulation or a Delaunay triangulation from a digital terrain model. It refines the triangulation iteratively to the desired precision. Finally, a parallel algorithm for 3-D Delaunay triangulation has been designed, and techniques have been developed to cope with issues such as load-balancing and robustness in the application of this algorithm.

3 Invariant properties [3, 6]

Two studies of the uses of invariant properties in recognition were conducted. The first [6] dealt with a new class of invariant measures that can be used for robust texture classification; and the second [3] was concerned with invariant signatures of shapes that can be used for efficient retrieval from an image database.

A new method of texture analysis and classification has been developed based on a local center-symmetric covariance analysis, using Kullback (log-likelihood) discrimination of sample and prototype distributions. The analysis has led to the development of generalized, invariant, local measures of texture having center-symmetric patterns, a property which is characteristic of most natural and artificial textures. Two local center-symmetric auto-correlation measures have been introduced, including linear and rank-order versions (SAC and SRAC), together with related covariance measures (SCOV) and variance ratios (SVR). All of these measures are rotation-invariant, and three of them are locally gray-scale invariant. In classification experiments, their discriminant information has been compared to that of Laws' well-known convolutions, which have specific center-symmetric masks. It was found that the new covariance measures achieved very low classification error rates despite the fact that they employ abstract representations of texture pattern and grey-scale.

Image databases are conceptually much harder to deal with than conventional databases because the information they contain consists of images, rather than alphanumeric entities. Images of objects can be fuzzy and distorted, and they depend on the point of view from which the objects are seen. However, characteristics of the images which are invariant to changes in the viewpoint can be defined. These characteristics can be stored as "signatures" for the objects in an atlas database, thereby permitting efficient retrieval and matching, partial or total, regardless of viewpoint. Such invariant signatures are very useful because image matching is slow in high population image databases. The signatures can also be indexed easily using current database technology (e.g., the B-tree). Experiments with this class of invariant features have made use of a sample database consisting of images of different types of fruits. In these experiments, examples of queries in such an environment have been formulated, and strategies for query evaluation have been tested.

4 Image registration [11]

The theoretical component of the research on image registration involves the following question: Given any two views of some unknown textured opaque quadric surface in 3-D, is there a finite number of corresponding points across the two views that uniquely determine all other correspondences coming from points on the quadric? A constructive answer to this question is given, and this answer is then used to propose a so-called "nominal quadratic transformation" that can be used in practice to facilitate the process of achieving full point-to-point correspondence between two grey-level images of the same (arbitrary) object. The approach has been implemented and applied to a real image situation, as follows:

Figure 1a shows two images of a face taken from two distinct viewpoints. Achieving full correspondence between two views of a face is extremely challenging for two reasons. First, a face is a complex object that is not easily parameterized. Second, the texture of a typical face does not contain enough image structure for obtaining point-to-point correspondence in a reliable manner. There are a few points (on the order of 10–20) that can be reliably matched, such as the corners of the eye, mouth and eyebrows. These few points are used to determine the quadratic nominal transformation and the epipolar geometry. Optical flow techniques are then applied to "finish off" the correspondence in the remaining areas. The epipoles were recovered using an algorithm described by Faugeras, which employs a varying number of points. The results presented here used the minimal number of nine points, but a similar performance was obtained using a least squares solution based on more than nine points.

In Figure 1b, the left-hand image (which displays the overlayed edges extracted from the two images in Figure 1a) shows that typical displacements between corresponding points around the center region of the face are approximately 20 pixels. The right-hand image displays the overlayed edges for the affine transformed images; it shows that if an affine transformation is used to reduce the displacements near the center of the face, the displacements in peripheral regions are increased. The three points chosen were the two eyes and the right mouth corner. Notice that the displacement across the center region of the face was reduced, at the expense of the peripheral regions that were taken farther apart.

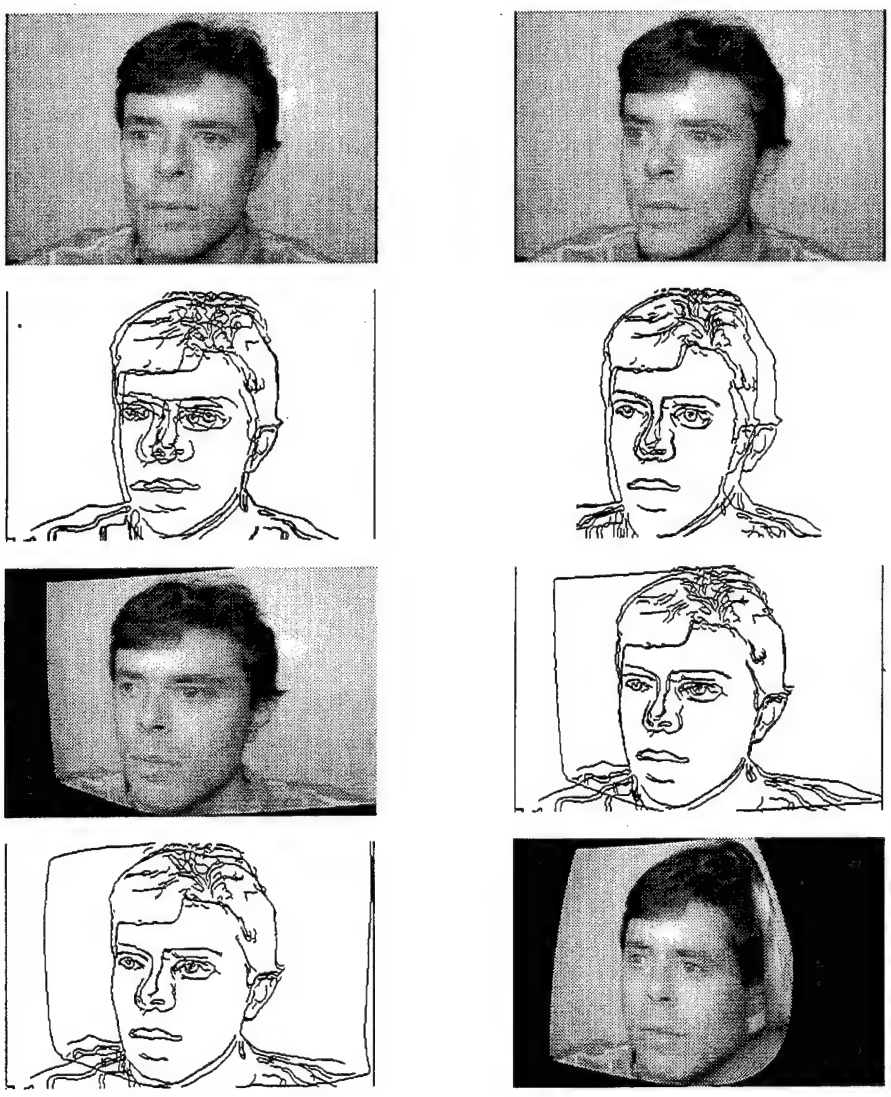


Figure 1: Quadratic transformation for image registration.

In Figure 1c, the left-hand image shows the result of applying the quadratic nominal transformation to the first view. The overlay of the edges of the second view, and the transformed first view, are shown in the right-hand image. It is seen that both the center of the face and the boundaries are brought closer together. Typical displacements have been reduced to approximately 1–2 pixels. The optical flow algorithm restricted along epipolar lines was applied between the transformed view and the second view. The final displacement field (nominal transformation due to the quadric, plus the residuals recovered by optical flow) was applied to the first view to yield a synthetic image that, if successful, should look much

like the second view. In order to test the similarity between the synthetic image and the second view, the overlay of the edges of the two images is shown in Figure 1d, left-hand image.

Finally, to illustrate that a quadric surface may yield unintuitive results, the right-hand image of Figure 1d shows the result of having a hyperboloid of two sheets as a quadric reference surface; note the mirror image on the right hand side. This is accidental, but evidently can happen with an unsuccessful choice of sample points.

5 3-D recovery [5, 10]

Two studies relating to the recovery of three-dimensional scene properties from single images were conducted. The first study [5] dealt with a new method of camera pose estimation by analysis of an image of a known object. The second study [10] developed an improved algorithm for deriving the three-dimensional shape of a surface from an analysis of the gray-level variations (shading) on an image of the surface.

A new method of computing the position and orientation of a camera with respect to a known object has been developed, using four or more *coplanar* feature points. Starting with the scaled orthographic projection approximation, this method iteratively refines up to two different pose estimates, and provides an associated quality measure for each pose. When the object's distance to the camera is large compared with the object's extent along the direction of the optical axis, or when the accuracy of feature point extraction is low because of image noise, the two quality measures are similar, and the two pose estimates are plausible interpretations of the available information. In contrast, known methods using a closed form pose solution for four coplanar points are not robust for distant objects in the presence of image noise because they provide only one of the two possible poses, and they may choose the wrong pose.

An improved shape from shading (SFS) algorithm, suitable for application to outdoor scenes, has also been developed. The algorithm obtains an accurate estimate of the azimuth of the illumination source. Accurate depth reconstruction is then achieved by using a new set of boundary conditions and adapting an improved technique for hierarchical implementation of the SFS algorithm. As a result, errors at the boundaries of images and in rotation of the reconstructed images have been corrected; such errors were characteristic of earlier methods. A typical result on a real image (of a face) will now be presented.

Figure 2 shows the (a) input image; (b) height map obtained by our SFS algorithm; (c) image synthesized from the reconstructed height map with the estimated parameters; (d) 3-D mesh plot of the height map obtained by SFS; (e) and (f) surface slant component maps obtained by the algorithm; (g)–(i) images synthesized from these slant maps, corresponding to illumination from directions opposite or orthogonal to the estimated direction

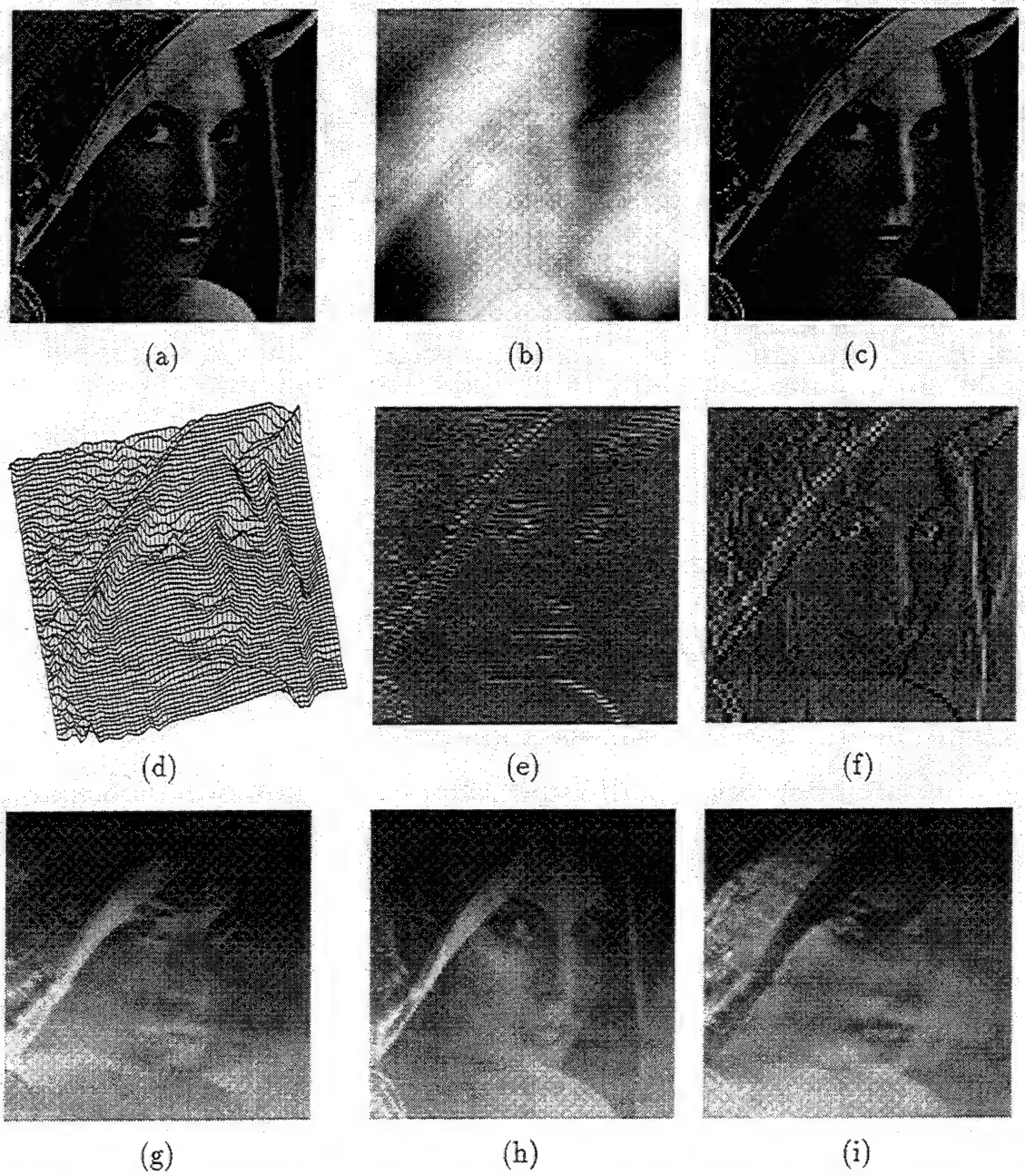


Figure 2: Shape from shading experiment on a face image.

for the input image. It can be seen from the figures that the shapes of the face and shoulder are recovered correctly and that features such as nose, lips, cheeks, chin, etc., are easily identified. Moreover, the images synthesized using different illuminant directions are consistent with the recovered shape information.

6 Motion analysis [8]

Extensive work has been done on the analysis of visual motion from long, noisy image sequences. The moving-camera/stationary scene and fixed-camera/moving object cases have been studied. The task involved is the estimation of motion and structure parameters from 2-D image coordinates. Under the central projection model, a kinematic model-based approach has been proposed to relate the time evolution of the parameters to the 2-D image plane feature coordinates. Two approaches, batch and recursive, have been developed to solve this highly nonlinear problem. In the batch approach, a cost function has been defined, and a conjugate gradient method has been applied to find all the parameters that minimize this cost function. In the recursive approach, an Iterated Extended Kalman Filter (IEKF) has been used to propagate and update the state variables.

Based on the assumption that the frame sampling rate is high enough, and thus, the motion is smooth over a short period of time, only the first order motion parameters have been used to model the camera (or object) motion. Therefore, in situations where departures from the assumed motion model exist, the batch approach provides rough estimates over the first few frames; while deviations from the assumed model are handled using the tracking ability of the Kalman filter. The standard rotation matrix is used to represent the rotational motion instead of the commonly used quaternions. This results in a simple linear plant model in the recursive method. Closed form solutions for the state and covariance transition equations are obtained without the use of the time-consuming numerical integration step.

In the monocular case, since motion and structure parameters are estimated at the same time, the processing of the batch algorithm over many frames would be time-consuming. The batch algorithm is therefore used to give rough estimates for all the parameters over the first few frames. These values can then be used as initial guesses for the recursive algorithm. In order to prevent the Kalman filter from diverging, a good initial guess for the covariance matrix is often necessary. This is done by computing the Cramér-Rao Lower Bounds for all the parameters, and using these bounds as the initial values of the covariance matrix.

In the binocular case, since the 3-D structures can be roughly estimated from the first pair of left and right images using the classic stereo triangulation method, both the batch

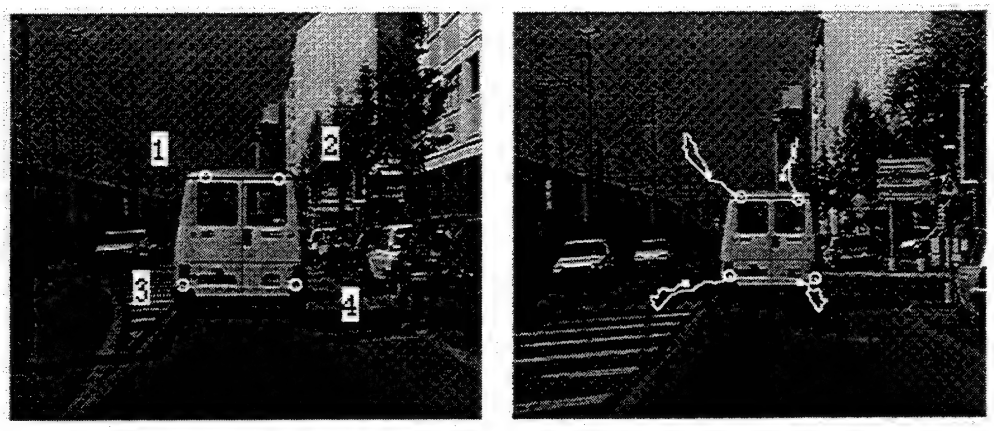
and recursive algorithms converge more reliably than in the monocular case. A proof of the uniqueness of the parameters has been given for the binocular case. It has been shown that three noncollinear feature points over three consecutive frames contain all the information needed for motion and structure estimation. Based on this proof, the rotational parameters can first be estimated using a deterministic method. Subsequently, all the other parameters can be estimated using a linear algorithm, leading to much faster implementation.

In real experiments, the accuracy of the algorithms depends heavily on the calibration of the camera being used. The problem of camera calibration has been addressed in a straightforward manner. If the 3-D structures of some feature points are available to us, then based on a simple batch algorithm, the image center and the field of view can be roughly estimated to improve the performance of the estimation process.

Several real image sequences have been carefully analyzed using these methods. The ground truth, when available, has been compared to the estimates. For most of the real sequences used, the inputs to the algorithms (2-D image coordinates) are all automatically detected and matched over frames. Despite the presence of various types of noise as well as the occlusion problem, the methods have yielded successful results, even though as few as four feature points were used.

As an example of an application of these methods, Figure 3a is the first frame of an image sequence showing a van driving in traffic. Feature points were automatically detected and tracked in all the frames; their locations in the first frame are marked by white circles in Figure 3a. Figure 3b shows the feature points in the 45^{th} frame; for comparison, their locations in the first frame are marked by white squares, and their trajectories in the 45-frame sequence are shown as white curves. For this sequence, ground truth information as well as camera calibration parameters were not available, so the field of view was arbitrarily assumed to be 24 degrees and the center of the image (optical center) to be coincident with the center of each image. Six-frame batch estimates of the structure and motion parameters are shown in Tables 1 and 2.

In the sequence, there is almost no rotational motion. The distance between the camera and the van in front decreases until frame 14 (the relative velocity is almost zero at this time). Subsequently, the van starts to accelerate faster than the camera, so the distance



(a) (b)

Figure 3: Image plane trajectories of feature points in the Van Sequence: (a) first frame; (b) 45^{th} frame.

begins to increase as though the camera were moving in the negative direction.

Table 1: Structure estimates for the Van Sequence (batch method).

Feature points	Estimated structure		
1	0.045517	-0.038920	0.998763
2	-0.036096	-0.036276	0.999079
3	0.069840	0.060836	1.001770
4	-0.056859	0.060727	1.0

Table 2: Motion parameter estimates for the Van Sequence (batch method).

Estimated rot. center	Estimated rot. velocity	Estimated trans. velocity
0.01002 0.00962 0.0	0.00279 0.00108 0.00140	-0.00078 0.00321 0.03766

These qualitative observations are very consistent with the translational velocity estimated by the Kalman filter using the six-frame batch output as its input. The result is shown in Figure 4. This experiment brings to light an advantage in the use of a Kalman filter in motion estimation based on long frame sequences: its adaptability to situations where we have a higher-order motion than that assumed by the model.

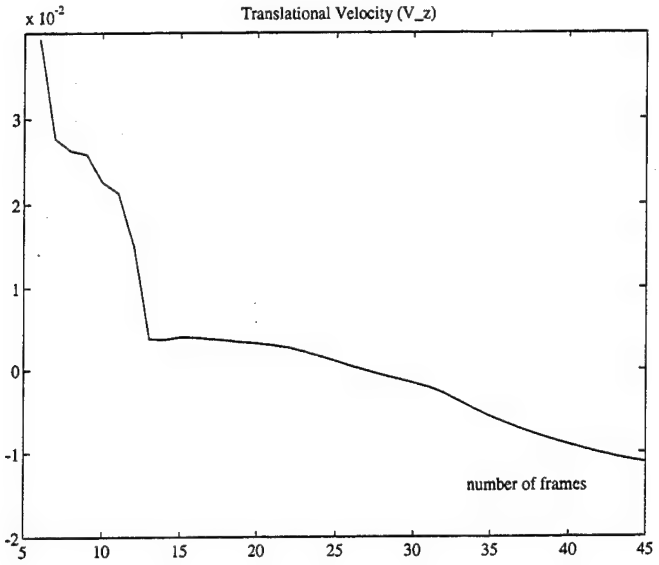


Figure 4: Translational velocity for the Van Sequence computed by the Kalman filter.

7 Vision-based navigation [2, 4]

Research on navigational applications of vision has dealt with two topics: The use of neural networks for autonomous road following [2], and the integration of visual information into robot control systems.

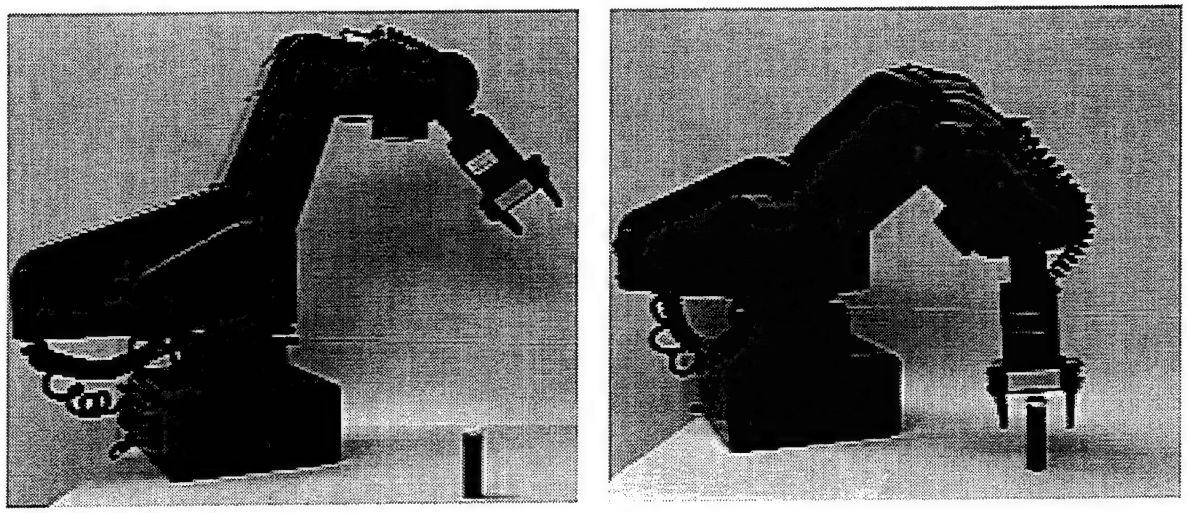
Recently, connectionist architecture approaches have been used to solve the autonomous visual road-following problem. Carnegie-Mellon University, using its Navlab vehicle, implemented a feed-forward multilayer perceptron (FMLP) network controller called ALVINN that very successfully performed visual road-following. ALVINN, however, experienced problems with primitive road feature retention, spurious control signals in the presence of structured noise, and learning anomalous driving situations. At the University of Maryland, an alternative connectionist architecture, called a Radial Basis Function (RBF) network, has been successfully used for visual road-following. A controller based on each network architecture was built, and the performance of the two controllers was evaluated using a driving simulator. The FMLP controller experienced the same problems on the driving simulator that ALVINN experienced in real road-following. The RBF controller did not experience any of the problems that the FMLP network experienced, and did not suffer any negative side-effects. The RPF network has been successfully used to drive the Carnegie-Mellon vehicle in highway traffic.

Traditionally, a robot's visual system is assigned the task of reconstructing the shape of the surrounding scene in the form of a depth map, which can be exploited to solve navigation problems by means of trajectory planning, control of mechanisms, etc. Unfortunately, as an overview of the state of the art in visual reconstruction reveals, it is still impossible to reliably compute depth maps, due to the fact that all shape-from-x problems are mathematically ill-posed (they have no unique solution and/or the solution is unstable). Furthermore, judging by progress made in recent years in problems such as road following and visual servoing, it appears that the depth map may, after all, not be the most suitable data representation for such tasks.

An image-based approach to the navigation problem has been developed in which visual processes are closely and actively integrated with robot control. It has been shown that task-

specific visual information with a minimum of structure is sufficient to accomplish classical navigational tasks such as obstacle avoidance and hand/eye coordination. Through these examples, it has been demonstrated that the use of vision actually simplifies the solution of robotics problems, allowing real-time control without the complex calibration procedures traditionally required.

It has been assumed in the past that the robot's trajectory is determined exclusively by the location of the goal and that the appropriate visual data can be acquired as the robot progresses toward the goal. In particular, the trajectory planning module never takes into account the needs of the visual module. It has been recognized that the stability and accuracy of visual algorithms are affected by the motion of the observer, but not much work has been done on deciding what constitutes a "good" motion. A quantitative criterion has been proposed for the evaluation of the goodness of a particular action, and it has been shown how this additional layer of planning can be incorporated into the low-level control strategy, together with higher-level symbolic reasoning.



(a) (b)

Figure 5: A grasping task: initial (a) and desired final (b) images.

These ideas have been demonstrated experimentally in connection with the vision-based control of a grasping task performed by a robot arm. Figure 5a shows the arm in its initial configuration and the object (dark cylinder) that must be grasped, as seen by the controller. Posed in terms of the visual input, the grasping problem is for the controller to "act" on the

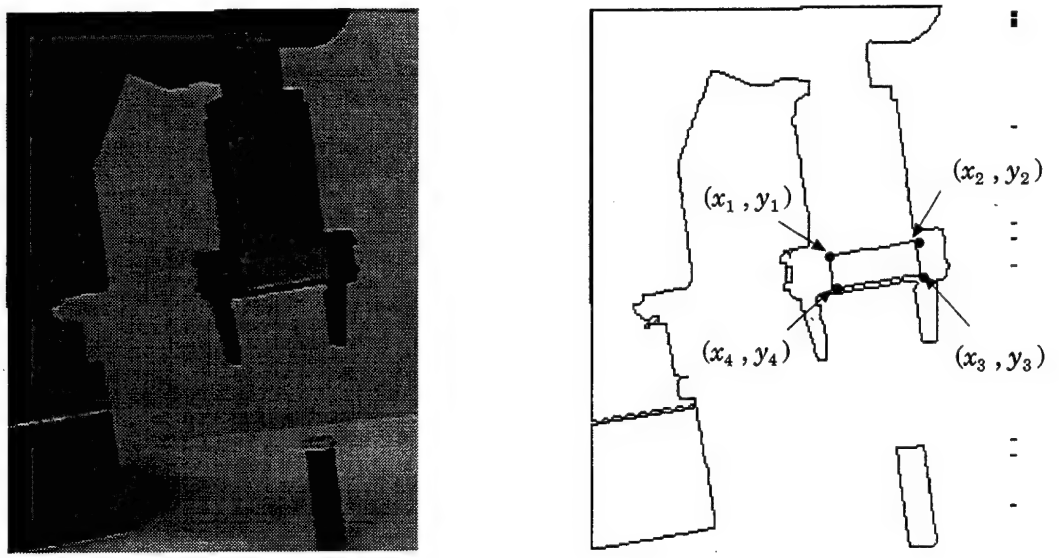


Figure 6: Selection of image features for grasping task.

image, to move the manipulator's joints so the scene appears as in Figure 5b. The joint space of the manipulator is mapped directly into a space of image features (the "camera space"); typically, these features are the coordinates of distinctive points in the image, as illustrated in Figure 6. The joint movements which lead to the desired final position of the hand are graphically illustrated in Figure 7 for a two-degree-of-freedom manipulator; the initial and final hand positions, and the trajectory of the hand, are plotted (as white points joined by a black curve) on a "perceptual control surface" (PCS). In effect, this approach does not require previous calibration of the system. It uses the image features directly to control the arm, rather than first using the images to compute the pose of the arm and then controlling the movement of the arm from one pose to another.

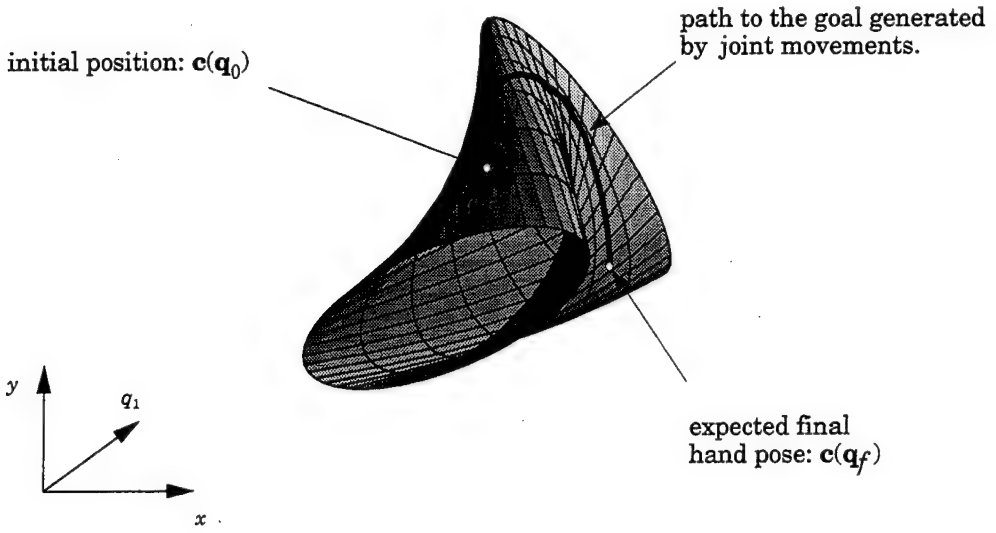


Figure 7: Trajectory planning on the perceptual control surface.

8 Function-based recognition [12]

An approach to function-based object recognition has been developed that reasons about the functionality of an object's primitive parts. The popular "recognition by parts" shape recognition framework has been extended to support "recognition by functional parts", by combining a set of functional primitives and their relations, with a set of abstract volumetric shape primitives and their relations. Previous approaches have relied on more global object features, often ignoring the problem of object segmentation; as a result, these approaches are generally restricted to range images of unoccluded scenes. However, the shape primitives and relations can be easily recovered from superquadric ellipsoids which, in turn, can be recovered from either range or intensity images of occluded scenes. Furthermore, the proposed framework supports both unexpected (bottom-up) object recognition and expected (top-down) object recognition. The approach has been demonstrated on a simple domain by recognizing a restricted class of hand-tools from 2-D images.

As an illustration of the approach, Figure 8a shows an image of a mallet, and Figure 8b shows the segmented region image. Without any a priori knowledge of scene content, each of the functional primitives, namely the end-effector and handle, is deemed equally likely to appear in the image. The algorithm arbitrarily chooses the end-effector (mallet head) and maps it to a search for a blob in the image. The algorithm rank-orders regions in the image according to their ratio of area to extent (computed from the bounding box). The large region is chosen first and the bottom-up algorithm is used to recover the most likely interpretation of the region and its neighbors. The two most likely recovered volumes are shown in Figures 8c–d, corresponding to the head and handle of the mallet, respectively. Those portions of the bounding contour used to infer part identity are highlighted in the image. Superquadrics are fitted to each part, as shown in Figures 8e–f. Intermediate grey values along contour portions represent locations of image forces acting on the superquad. Due to the absence of forces at the junction between the two parts (no contours at the junction end of the handle), the fitted handle was not "pulled" all the way to the junction. Using the recovered superquadric parameters, the parts were classified qualitatively as a blob and a stick, respectively. Since the search procedure is looking for the mallet head (end-

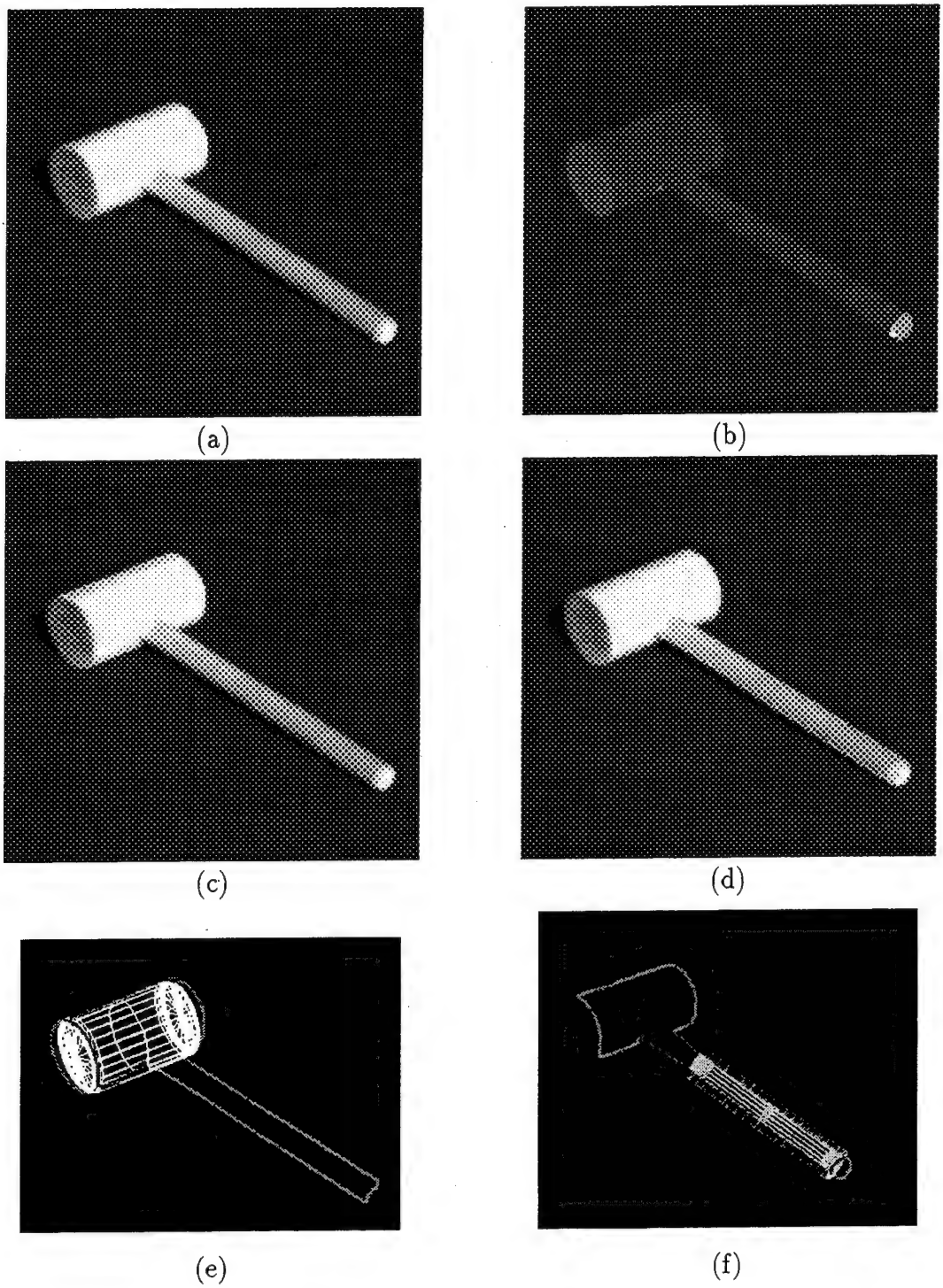
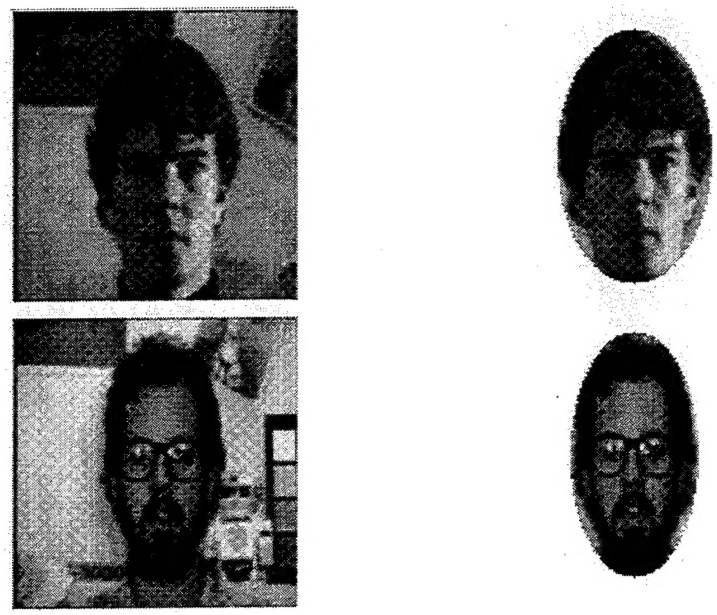


Figure 8: Recognition by functional parts.

effector), it chooses the blob, and proceeds to search for the handle in the vicinity of the recovered blob. Due to region undersegmentation, the regions, corresponding to the body surfaces of the head and handle of the mallet, are joined. However, those contours not used to recover the head, but still belonging to the large region, are free to be part of other recovered volumes. Since a stick has already been recovered, and its defining contours were not used to infer the blob, the handle can be instantiated in the image. The last step in recognizing the object is to satisfy the functional relation between the two parts; this relation is mapped into a spatial constraint on the part junction. Since the computed relative orientation of the two parts is such that their axes are orthogonal, and since the junction occurs at the end of the handle and at the middle of the head, the algorithm successfully verifies the hammer in the image.

9 Face recognition [9]

Techniques have been developed for segmentation and identification of human faces from grey scale images with clutter. The segmentation utilizes the elliptical structure of the human head. It uses the information present in the edge map of the image and through some preprocessing, separates the head from the background clutter. An ellipse is then fitted to mark the boundary between the head region and the background. The identification procedure finds feature points in the segmented face through a Gabor wavelet decomposition and performs graph matching. The segmentation and identification algorithms have been tested on a database of 48 images, showing 16 persons, with encouraging results. Examples of good results of the segmentation process are shown in Figure 9.



(a) (b)

Figure 9: Good results of face segmentation. (a) Input image; (b) extracted image.

10 Document understanding [1]

Many document image understanding problems require a more comprehensive examination of document features than is typically deemed necessary for recognition tasks. Specifically, these problems require a detailed analysis of stroke and sub-stroke features in the document image with the goal of obtaining information about the environment, or process which created the document, and establishing a context for understanding.

Document image understanding research at the University of Maryland has introduced the concept of *recovery* into the document domain. A "stroke platform" representation has been developed which establishes a verifiable "link to the pixels", and its usefulness for recovery tasks has been demonstrated. This representation makes it possible to overcome many of the problems associated with the rapid, irreversible abstraction associated with traditional document processing methods and provides the basic framework for an analysis of handwritten documents. By obtaining a detailed description of the document and its properties, it is possible to establish a context for analysis and to validate assumptions about the domain. Research has been conducted on this project on several document image understanding problems:

- Demonstrating the successful use of the stroke platform for the problem of interpreting and reconstructing junctions and endpoints,
- Exploring the effects of the handwriting process on the document by the development of a model for instrument grasp and a study of its effects on pressure features,
- Posing and providing an approach to the problem of recovering temporal information from static images of handwriting,
- Addressing various sub-tasks arising in the problem of processing form documents, and
- Extending the detailed analysis philosophy to demonstrate its feasibility in related document domains.

11 Bibliography of reports under this contract

1. D.S. Doermann, "Document Image Understanding: Integrating Recovery and Interpretation." CAR-TR-662, CS-TR-3056, April 1993.

2. M. Rosenblum and L.S. Davis, "The Use of a Radial Basis Function Network for Visual Autonomous Road Following." CAR-TR-666, CS-TR-3062, May 1993.

3. I. Weiss, W.G. Aref, E. Rivlin and H. Samet, "Geometric Invariants for Image Databases." CAR-TR-667, CS-TR-3063, May 1993.

4. J.-Y. Hervé, "Navigational Vision." CAR-TR-669, CS-TR-3065, May 1993.

5. D. Oberkampf, D.F. DeMenthon and L.S. Davis, "Iterative Pose Estimation Using Coplanar Feature Points." CAR-TR-677, CS-TR-3098, July 1993.

6. D. Harwood, T. Ojala, M. Pietikäinen, S. Kelman and L.S. Davis, "Texture Classification by Center-Symmetric Auto-Correlation, Using Kullback Discrimination of Distributions." CAR-TR-678, CS-TR-3099, July 1993.

7. Y.A. Teng, "Parallel Processing of Geometric Structures: Visibility and Triangulation Algorithms." CAR-TR-680, CS-TR-3120, August 1993.

8. T.-H. Wu, "Estimation of Motion and Structure from Long Noisy Image Sequences." CAR-TR-689, CS-TR-3146, October 1993.

9. S.A. Sirohey, "Human Face Segmentation and Identification." CAR-TR-695, CS-TR-3176, November 1993.

10. H. Singh and R. Chellappa, "An Improved Shape from Shading Algorithm." CAR-TR-700, CS-TR-3218, February 1994.

11. A. Shashua and S. Toelg, "The Quadric Reference Surface: Applications in Registering Views of Complex 3-D Objects." CAR-TR-702, CS-TR-3220, February 1994.

12. E. Rivlin, S.J. Dickinson and A. Rosenfeld, "Recognition by Functional Parts." CAR-TR-703, CS-TR-3222, February 1994.